

# Reconstruction of vocal tract geometries from biomechanical simulations

Saeed Dabbaghchian<sup>1</sup>  | Marc Arnela<sup>2</sup> | Olov Engwall<sup>1</sup> | Oriol Guasch<sup>2</sup>

<sup>1</sup>Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

<sup>2</sup>GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Spain

## Correspondence

Saeed Dabbaghchian, Department of Speech, Music, and Hearing, KTH Royal Institute of Technology, Stockholm SE-10044, Sweden.  
Email: saeedd@kth.se

## Funding information

Agencia Estatal de Investigación and FEDER, EU, Grant/Award Number: TEC2016-81107-P; Seventh Framework Programme, Grant/Award Number: 308874

## Abstract

Medical imaging techniques are usually utilized to acquire the vocal tract geometry in 3D, which may then be used, eg, for acoustic/fluid simulation. As an alternative, such a geometry may also be acquired from a biomechanical simulation, which allows to alter the anatomy and/or articulation to study a variety of configurations. In a biomechanical model, each physical structure is described by its geometry and its properties (such as mass, stiffness, and muscles). In such a model, the vocal tract itself does not have an explicit representation, since it is a cavity rather than a physical structure. Instead, its geometry is defined implicitly by all the structures surrounding the cavity, and such an implicit representation may not be suitable for visualization or for acoustic/fluid simulation. In this work, we propose a method to reconstruct the vocal tract geometry at each time step during the biomechanical simulation. Complexity of the problem, which arises from model alignment artifacts, is addressed by the proposed method. In addition to the main cavity, other small cavities, including the piriform fossa, the sublingual cavity, and the interdental space, can be reconstructed. These cavities may appear or disappear by the position of the larynx, the mandible, and the tongue. To illustrate our method, various static and temporal geometries of the vocal tract are reconstructed and visualized. As a proof of concept, the reconstructed geometries of three cardinal vowels are further used in an acoustic simulation, and the corresponding transfer functions are derived.

## KEYWORDS

acoustic model, biomechanical model, speech production, vocal tract geometry

## 1 | INTRODUCTION

The vocal tract plays a role in speech production, respiration, and swallowing, and understanding its anatomy and physiology may thus be beneficial, eg, for treatment of related disorders. However, the vocal tract is not directly visible, which is a barrier for such studies, and sophisticated medical imaging techniques, such as magnetic resonance imaging (MRI) or computed tomography (CT), are required to determine the geometry of the vocal tract. Once the images have been acquired, a dedicated boundary detection procedure is employed to segment the vocal tract, and to create a 3D geometry,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2018 The Authors. *International Journal for Numerical Methods in Biomedical Engineering* published by John Wiley & Sons, Ltd.

which may then be used for numerical simulations. This approach has been successfully applied in previous works to study the production of speech sounds (see, eg, previous literatures<sup>1-7</sup>) and simulation of fluid in sleep apnea (see, eg, previous literatures<sup>8-14</sup>). All these works follow a two-step procedure: acquiring the vocal tract geometry to define a computational domain and conducting acoustic/fluid simulations. The first step requires a human subject to acquire images of his/her vocal tract and to reconstruct its geometry.

A different approach consists in determining the vocal tract geometry from a model, which has the advantage that alternative anatomical or articulatory configurations may be tested without acquiring new medical data. Direct imaging of the vocal tract is absolutely essential for basic knowledge of anatomy and articulation, but supplementing it with vocal tract modeling may be beneficial for the research in the field, as it allows for more flexibility and overcomes some of the drawbacks of medical imaging. To list a few, the amount of manual work to create a 3D static geometry from images is significant; MRI cannot capture the teeth; scanning time is usually long, especially when a high spatial resolution is needed; because of the scanning time, it is difficult to capture the deformation of the vocal tract with sufficient spatiotemporal resolution (see Fu et al<sup>15</sup> and Lingala et al,<sup>16</sup> for the most recent achievement in real-time MRI); for each new articulation or subject, one needs to acquire new images and repeat the whole procedure. On the other hand, if using a biomechanical model, it is possible to accurately simulate the movement of the dynamical structures and hence determine the vocal tract shape at each time step. The temporal resolution is limited only by the time step of the simulation, and the spatial resolution of constrictions (eg, between the tongue and the hard palate) is high, whereas the contrast between air and tissue may be low for narrow constrictions using MRI. In addition to these advantages, a biomechanical model is sometimes the exclusive choice to approach and investigate problems, such as predicting the consequence of a glossectomy,<sup>17,18</sup> hypothesizing English [r] variants,<sup>19</sup> or across-subject variability of the muscle activations during speech production.<sup>20</sup> Simulation of swallowing may be another application of such an approach (see, eg, previous literatures<sup>21-25</sup>).

Therefore, it is clear that reconstructing the vocal tract geometry from a biomechanical model has important applications. However, whereas there are well-established methodologies and tools to reconstruct static vocal tract geometries from medical images, little attention has been paid to acquire the vocal tract geometry from a biomechanical simulation. This is moreover a rather complex task. To clarify why, we need to review the process of building a biomechanical model. It starts with imaging of the anatomical structures (the tongue, the palate, etc) in a neutral position. Each physical structure such as the tongue and the mandible is segmented in the images and converted to a 3D geometry (either a volume or surface mesh). The geometry of the muscle fibers, constraints such as attachment points between soft and hard structures, and other properties such as mass and stiffness are defined. Then a set of physics equations is solved numerically to predict the behavior of the system, which results in shaping the vocal tract boundary. A key point to notice here is that the vocal tract is not a physical structure, but a cavity enclosed by several other structures, and hence its geometry (ie, boundary) is only implicitly defined, by the geometries of the surrounding structures. Obtaining the vocal tract cavity would be trivial if all surrounding geometries had a clearly aligned set of vertices/edges/faces, as we will show in Section 3. However, this condition is generally not satisfied in biomechanical speech production modeling. Existing gaps and overlaps between the structures of the biomechanical model would cause holes in the vocal tract cavity mesh, making it unsuitable for acoustic/fluid simulations. These artifacts are unavoidable because of the process involved in geometry extraction from medical images or missing small structures in the biomechanical model. Furthermore, such gaps and overlaps may also appear again when the structures start to move.

In early studies, because of the imaging technology available at that time, it was not possible to acquire multislice images simultaneously. Consequently, vocal tract models were mainly limited to the midsagittal plane (eg, Maeda<sup>26</sup>). However, as the midsagittal contour of the vocal tract is not in itself suitable to calculate the formant frequencies or to synthesize the sound, the vocal tract area function has to be determined using its midsagittal contour.<sup>27,28</sup> Despite the drawbacks, including accuracy (see Soquet et al,<sup>29</sup> for a review), it was the best available approach at that time and was used in many studies, including the 2D biomechanical model by Payan and Perrier.<sup>30</sup> Later on, simultaneous multislice imaging made it possible to measure the area function directly using 3D images<sup>31-34</sup> and also to develop 3D biomechanical models<sup>35-41</sup> that are more realistic regarding the anatomy (eg, fiber direction of the muscles) and mechanical properties (eg, muscular hydrostat) of the structures. Despite the progress in biomechanical modeling, little attention yet has been paid to reconstructing the complex vocal tract geometry in 3D,<sup>42</sup> and the area function is still the most common representation.

As the main contribution of this work, we propose a method that automatically reconstructs the 3D realistic vocal tract geometry from a biomechanical simulation. In an early attempt by the authors,<sup>43</sup> a method was proposed to construct a tube-like cavity excluding subbranches, which was used for the numerical simulation of three cardinal vowels. Later, the method was revised substantially<sup>44</sup> to improve the stability and speed for the numerical simulation of vowel-vowel sequences. Still, it was limited to the main cavity. In this work, a new and more accurate approach based on the growing

circles (GCs) concept is proposed to reconstruct a detailed geometry of the vocal tract including the piriform fossa, the interdental space, and the sublingual cavity. This method may be applied to any biomechanical or geometrical model to visualize the vocal tract shape and to link the model to acoustic/fluid simulations. Using an adapted version of the FRANK biomechanical model,<sup>41</sup> we demonstrate how different examples of the vocal tract geometry are reconstructed, including the geometries corresponding to vowels and a lateral approximant. In addition to static geometries, the method is capable of generating deformable geometries (ie, a vocal tract with moving boundaries). As a proof of concept, the reconstructed vocal tract cavities have been used in this work for the 3D finite element production of vowels. As far as the authors know, it is the first time that a link is established between a 3D biomechanical model and a 3D numerical simulation of voice. This offers a great potential for future applications of 3D voice production, especially those involving dynamic sounds like vowel-vowel utterances or syllables, due to the MRI data limitations mentioned above. The outcome of this work also represents a further step in the ambitious field of virtual human physiology, which is expected to play a predominant role in medical surgery and treatments in a not so distant future.

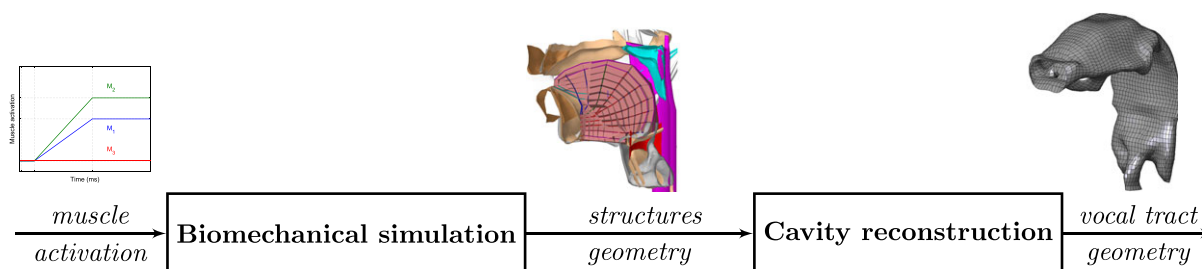
The proposed method has been developed to be used in a 3D biomechanical model, but it may be applicable to other similar problems, in particular boundary detection in medical images. There are numerous well-established tools for the latter problem, working on pixel value thresholds in grayscale images. However, in some cases where the contrast between air and tissue is low, the proposed method could contribute to better automatic definition of the vocal tract geometry by avoiding common artifacts (cf Section 3 for discussion of such artifacts). It should further be noted that although there are similarities between the proposed method and existing ones for medical images, a specific method like the one presented here is required for biomechanical modeling of the speech apparatus, due to the properties of the model (cf Section 3).

The rest of the paper is structured as follows. Section 2 describes the biomechanical simulation, including the input of muscle contractions that move and shape the articulators. Section 3 explains the cavity reconstruction method, in particular how it detects the boundaries of the vocal tract by processing the geometries of the surrounding anatomical structures. Section 4 presents the resulting vocal tract geometries for a selection of phonemes. Section 5 illustrates how the extracted vocal tract geometries may be used for acoustic simulations of the 3D wave equation in the time domain using the finite element method (FEM). It should be noted that the purpose of these simulations is not to perform a thorough investigation of the acoustic properties of the vocal tract (this has been done previously using 3D geometries from medical images and will be studied in detail for the biomechanical model in a follow-up study) but to demonstrate that the extracted vocal tract geometries are suitable for 3D acoustic simulations. Section 6 concludes the paper.

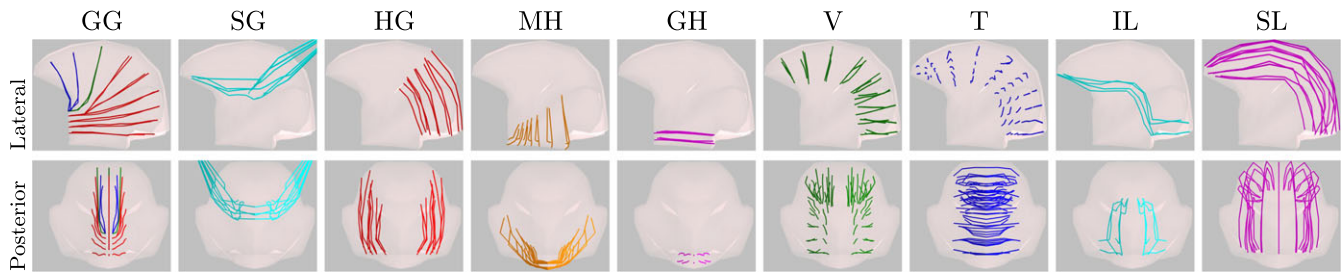
## 2 | BIOMECHANICAL SIMULATION

Figure 1 depicts the block diagram of the vocal tract geometry reconstruction from a biomechanical simulation. A temporal signal of muscle contraction is input to the biomechanical simulation, which deforms and moves the structures such as the tongue and the jaw. At each time step, the geometries of the structures are available and considered as the output of the biomechanical simulation. However, as the biomechanical simulation does not provide the vocal tract geometry since it is a cavity bounded by other structures, the cavity reconstruction method is then used to acquire its geometry at each time step as shown in Figure 1. This explicit representation of the vocal tract geometry offers the possibility to use existing numerical methods for acoustic/fluid simulation. It also provides an accessible visualization of the vocal tract geometry.

The biomechanical model used in this work is an adapted version of FRANK<sup>41</sup> developed in ArtiSynth.<sup>45</sup> The FRANK model, for which the midsagittal view is shown in Figure 1, is composed of rigid bodies, including maxilla, mandible,



**FIGURE 1** Reconstruction of the vocal tract geometry from a biomechanical simulation



**FIGURE 2** Lateral and posterior views of tongue muscles' fiber direction. GG: genioglossus, divided into three parts—posterior (GGP in red), middle (GGM in green), and anterior (GGA in blue); SG: styloglossus; HG: hyoglossus; MH: mylohyoid; GH: geniohyoid; V: verticalis; T: transversus; IL: inferior longitudinal; SL: superior longitudinal

hyoid bone, thyroid, and the deformable bodies face, tongue, soft palate (velum), pharynx wall, and larynx structure. In this work, the original model was modified in order to limit the computational cost, to improve simulation stability, and to give realistic acoustic simulation results. It should be noted that the different anatomical structures in FRANK are partly built on different subjects and that modifications are therefore required to generate adequate acoustic results. Adapting the entire model to an individual speaker is beyond the scope of this work (see Harandi et al,<sup>46</sup> for subject-specific modeling), and we instead make smaller adjustments that are necessary for our purposes. To speed up the simulation and improve the stability, the face mesh was removed but the geometry of the lips was kept. The downside is that we lose the control of the lip rounding, since the relevant muscles have been removed. In our model, the upper lip is attached to the maxilla and is static, while the lower lip is attached to the mandible and follows its movement. An internal surface of the face mesh was also cropped and modified to cover the cheeks and the ramus of the mandible. The thyrohyoid membrane (a thin layer between the thyroid and hyoid bones) and the mouth floor are missing in FRANK. Using anatomical knowledge, they were added as skin meshes. Following our previous work,<sup>43</sup> the lower part of the pharynx geometry (in the laryngopharynx region) was modified, and the larynx structure was translated 7 mm towards the anterior, and the uvula was slightly altered compared with the original FRANK model. The reason was to reach the typical formant frequencies for the three cardinal vowels and to achieve a normal shape and size of the piriform fossa. These modifications cause partial overlap between anatomical structures in the laryngopharynx region. The epiglottis was removed from the model, since we lack data on how to control it systematically. Soft palate muscles were activated to close the nasal cavity in all simulations. Again, to speed up the simulation and improve the stability, only the tongue is modeled as a deformable body, and all other structures are treated as rigid bodies, of which the lower lip, mandible, and hyoid bone are dynamic while the others are static. In all simulations, contact is modeled between the tongue and the maxilla, the mandible, the pharynx, and the soft palate. The tongue is controlled by the 11 muscles shown in Figure 2, ie, genioglossus posterior (GGP), genioglossus middle (GGM), genioglossus anterior (GGA), styloglossus (SG), hyoglossus (HG), geniohyoid (GH), mylohyoid (MH), verticalis (V), transversus (T), inferior longitudinal (IL), and superior longitudinal (SL). Two muscle groups control the jaw opening/closing. For more details about the biomechanical model, we refer the reader to the previous publications on ArtiSynth.<sup>41,47</sup>

### 3 | CAVITY RECONSTRUCTION

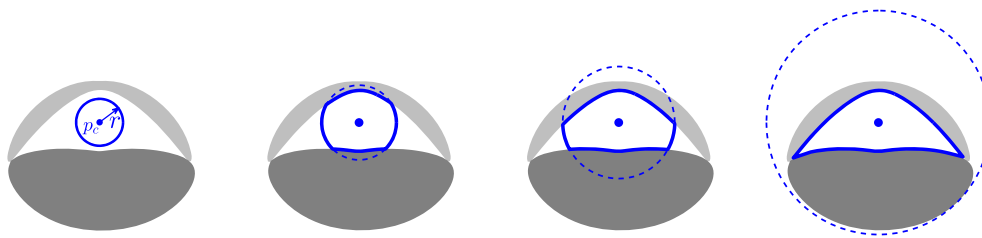
We define a cavity as a 3D region that is enclosed by two or more objects. The geometry of such a cavity is implicitly expressed by the boundaries of the surrounding objects. An explicit representation may be defined using geometrical Boolean (GB) operations (eg, Vatti<sup>48</sup>). This method works well when the surrounding boundaries are perfectly aligned. However, this is a strict requirement that is rarely satisfied in a biomechanical model. Instead, there are many gaps and overlaps between the boundaries of the surrounding objects, as explained above. Since these errors are unavoidable, we have developed a method that constructs the geometry of the cavity despite the presence of gaps and overlaps. Further, the vocal tract has an intricate morphology with branching, and subcavities may be formed by movement of the articulators. Developing a method to address the problem for arbitrary geometries is too complex and out of the scope of this work, but using anatomical knowledge of the vocal tract to constraint the problem whenever possible, we make the problem tractable. We aim to construct a geometry that is air-tight, manifold, self-intersection free, and has regular elements. Our proposed approach consists of finding cavity boundaries on several planar cross sections and then constructing a 3D geometry by triangulation between the cross sections. Section 3.1 elaborates on 2D boundary detection, and in

Section 3.2, we demonstrate how 2D boundary detection can be used in combination with slicing planes to detect 3D boundaries. Section 3.3 describes how the geometry of the cavity is constructed by addressing branching and triangulation problems of the 3D boundaries.

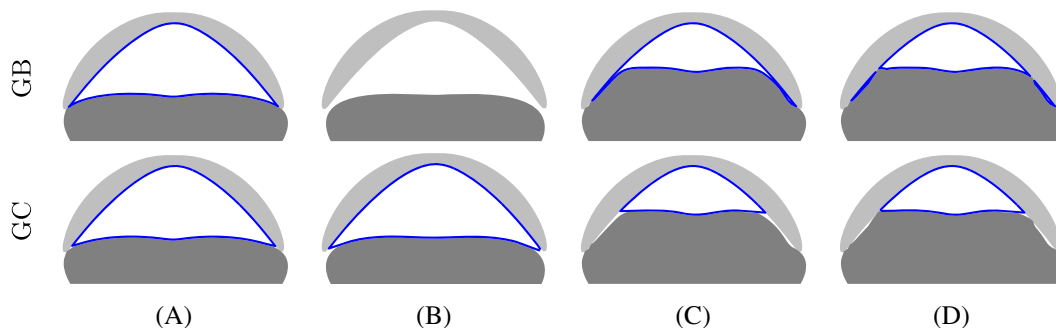
### 3.1 | 2D boundary detection

When a 2D region is enclosed by two or more polygons, the question arises as to how the boundaries of the region are determined. A similar problem arises in image segmentation, for which a popular approach is the region growing algorithm<sup>49</sup> that detects the boundaries of the region by growing an initial seed point. However, this algorithm works on images (ie, Cartesian grids), but it is not applicable in this work because of the need of manual seed selection and the described gaps/overlaps between anatomical structures. Instead, the idea of GCs has been developed in this work to address 2D boundary detection for models for which surrounding shapes are not guaranteed to form a well-defined boundary. Figure 3 illustrates the idea with an example. For the moment, we assume that the center of the enclosed region,  $p_c$ , is known. First a circle with radius  $r$  is positioned at point  $p_c$ . Then the circle starts to grow by increasing its radius  $r$  but limiting its growth by all surrounding boundaries, as shown in Figure 3 from left to right. More examples are shown in Figure 4 to illustrate how the GC method works in different situations in comparison with GB. In Figure 4A, adjacent boundaries of two polygons are perfectly aligned and both methods result in a similar output. However, when there is a gap, the GB does not identify any region. The size of the gap in Figure 4B is exaggerated for visual purposes, and only a tiny gap (on the order of numerical accuracy) is enough to get a null result. Figure 4C shows another boundary mismatch where the region is closed, but there is a narrow gap along the boundaries. This results in a region with a tail-like branch in the corners for GB. Another situation arises when there is a gap and overlap between two polygons as shown in Figure 4D. In this case, with GB, two extra sliver regions are identified in addition to the main region. Tail-like branches and slivers are considered to be artifacts and would be problematic in next steps.

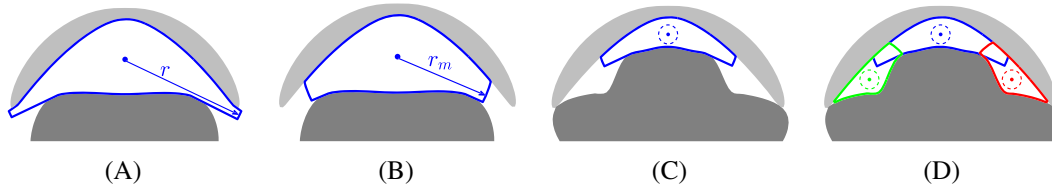
Although in all examples of Figure 4, the GC detects the boundaries of the enclosed region without generating any artifacts, there are still some other situations for which GC may not accomplish a correct result. In Figure 5A, there is a wide gap between two polygons, so the circle continues growing without any limitation. A maximum radius  $r_m$  is therefore defined for the GC, which then results in Figure 5B. Figure 5C shows another situation when a bulge prevents the circle from growing properly and the geometry of the enclosed region is not constructed correctly. Using three circles, instead of one, and the union of the results solves this issue, as shown in Figure 5D.



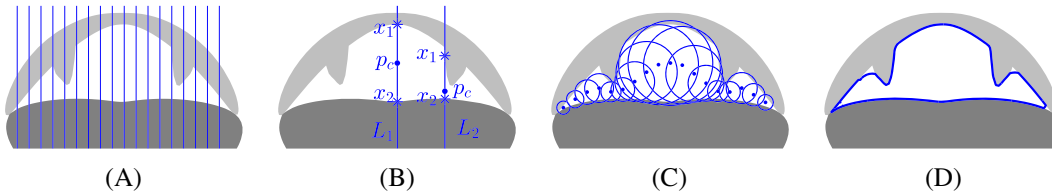
**FIGURE 3** The concept of growing circle: a circle grows from the center while it is limited by the surrounding boundaries



**FIGURE 4** 2D boundary detection using geometrical Boolean (GB) and growing circles (GCs) when (A) two polygons are perfectly aligned, (B) there is a small gap, (C) there is a narrow gap along the boundaries, and (D) there is a narrow gap and overlap



**FIGURE 5** Growing circle when there is a wide gap and bulge. (A) The circle continues to grow without limit when there is a wide gap. (B) The growing is limited by introducing the maximum radius  $r_m$ . (C) The circle fails to grow behind a bulge. (D) Using three circles instead of one solves the problem



**FIGURE 6** Calculation of the circles' center and maximum radius. (A) Scanning the region of interest by vertical lines. (B) Finding the circles' centers. (C) Resulting growing circles. (D) Geometry of the enclosed region

The examples in Figure 5 show how crucial it is to use the right parameters for the GC method including the number, center, and maximum radius of the circles. To have a general method that works for a relatively large set of enclosed regions, these parameters need to be determined adaptively. Figure 6 illustrates our approach. A set of vertical lines scans the whole region of interest as shown in Figure 6A (the reason for using vertical lines is the anatomical structures of the vocal tract, as we will see in Section 4). These lines intersect all boundaries, and the intersection points  $x_1$  and  $x_2$  are identified (see Figure 6B). A point on the line segment  $\overline{x_1x_2}$  that is equidistant from all intersected polygons is chosen as the center of the GC,  $p_c$ , as indicated in Figure 6B. In some cases, this is equivalent to finding the midpoint between  $x_1$  and  $x_2$  (see line  $L_1$  in Figure 6B). However, when the tangent of the boundary at one of the intersection points is close to a vertical line, the result will be different (see line  $L_2$  in Figure 6B). Once we have the centers, the maximum radius,  $r_m$ , is calculated as

$$r_m = K \times (d^2 + d_l^2)^{0.5} \quad (1)$$

with

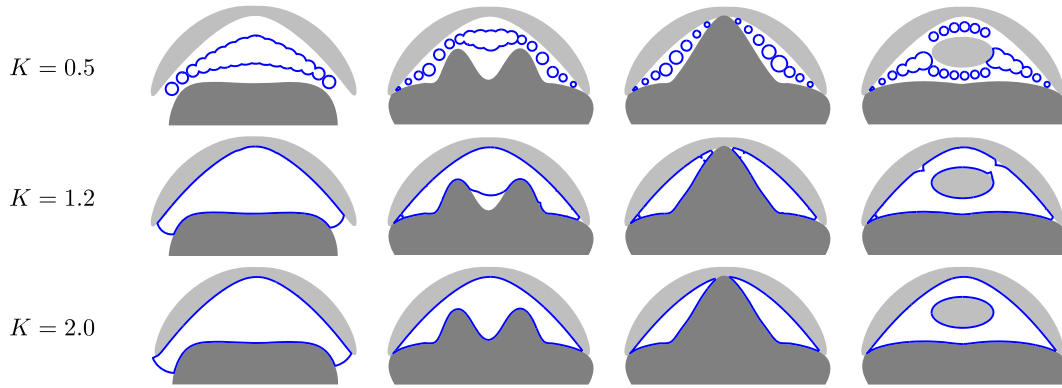
$$d = \min_{i \in \{1,2\}} \|x_i - p_c\|, \quad d_l = \frac{\|L_1 - L_2\|}{2}.$$

In this equation,  $\|\cdot\|$  is the Euclidean distance,  $d$  is the minimum distance between the center  $p_c$  and two intersection points  $x_1$  and  $x_2$ ,  $d_l$  is the distance between two consecutive vertical lines, and  $K$  is a constant. Using (1),  $r_m$  is calculated and Figure 6C shows the resulting GCs for this example. The geometry of the enclosed region is determined by unifying ( $\cup$ ) all the circles as shown in Figure 6D. The distance between the vertical lines affects the accuracy of the results, and it becomes even more important to detect the contact regions, which is a region where the surrounding boundaries meet each other. With large  $d_l$ , the algorithm may fail to detect a small contact region. If  $d$  is less than the threshold, then the line is considered to be in the contact region and hence no circle is determined in this case.

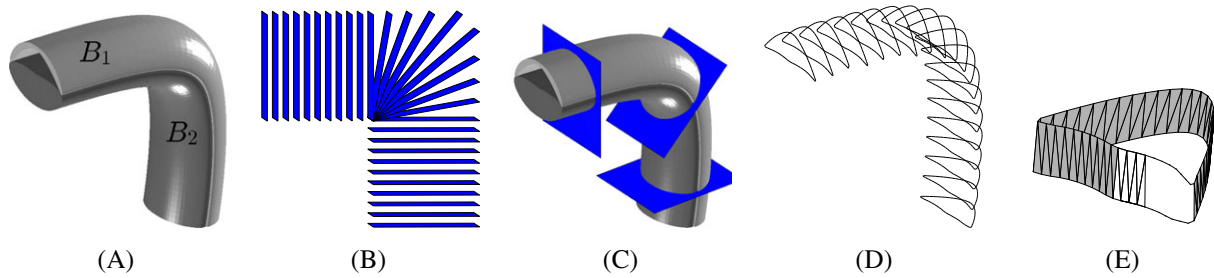
The constant value  $K$  should be chosen on the basis of the boundary shapes, as illustrated by the examples in Figure 7. Small  $K$  (usually less than 1) may result in several disconnected polygons. Large  $K$  (usually larger than 1) can guarantee a continuous polygon but may not accomplish good results when there is a wide gap. Furthermore, small  $K$  results in smoother shapes, and large  $K$  may result in a polygon with small angles.

### 3.2 | 3D boundary detection

In order to find the boundaries of a cavity (an enclosed region in 3D), the surrounding 3D geometries are sliced into planar contours. The GC method is then applied on each plane. On the basis of the vocal tract morphology, the slicing planes are arranged as a semipolar grid<sup>50</sup> composed of horizontal, oblique, and vertical sections (see Figure 8B). Such



**FIGURE 7** The influence of  $K$  on the resulting geometry



**FIGURE 8** (A) A vocal tract-like cavity. (B) Slicing planes arranged as semipolar grid. (C) Intersection of slicing planes with surrounding geometries. (D) Detected boundaries using growing circles. (E) Triangulation

slicing planes are approximately perpendicular to the midline of the cavity, and hence, we discretize the domain along its midline. Figure 8A shows a vocal tract-like cavity that is bounded by two geometries  $B_1$  and  $B_2$ . The slicing planes and their intersection with the 3D geometry are shown in Figure 8B and 8C. Figure 8D depicts the detected boundaries in each plane using the GC method.

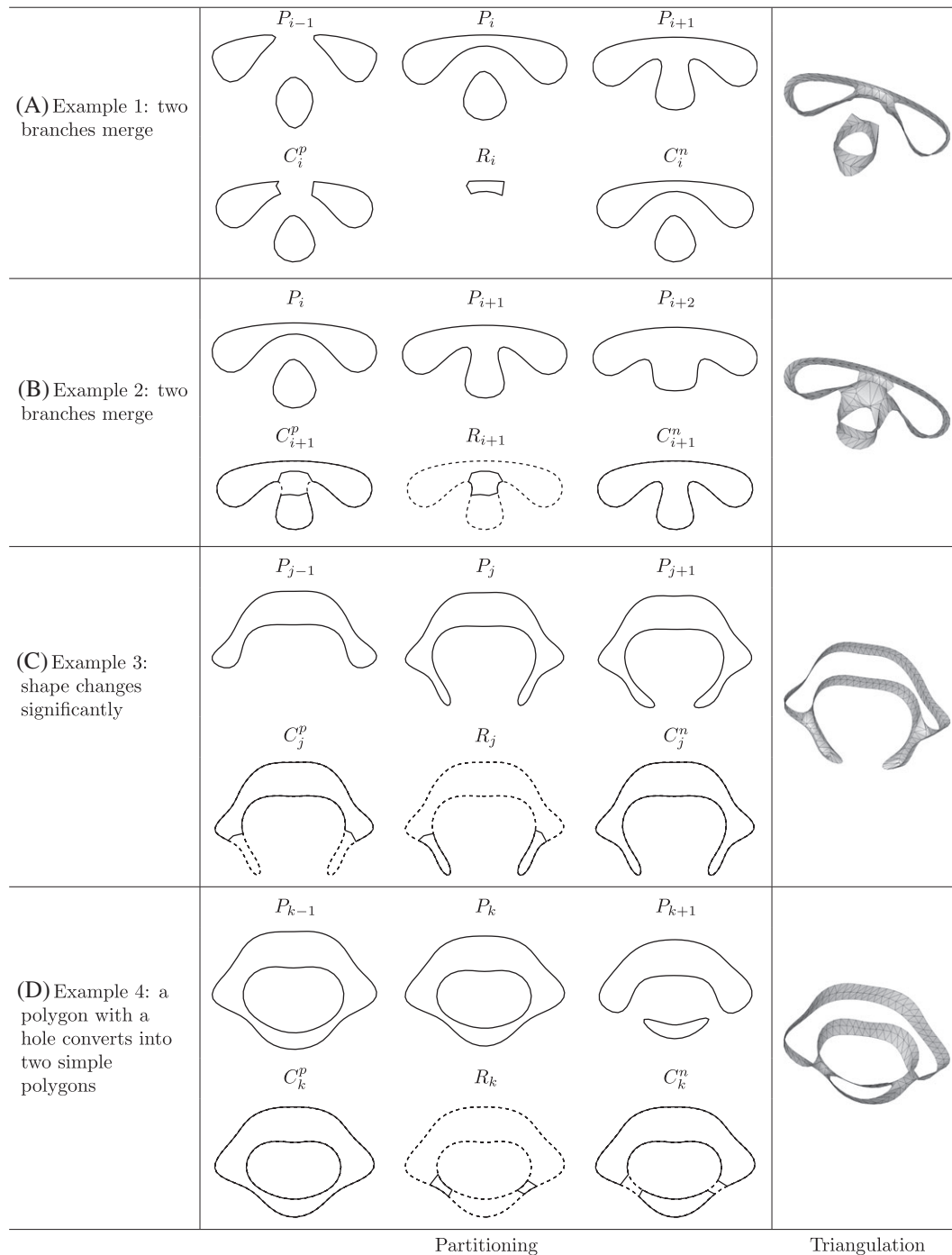
### 3.3 | Shape reconstruction

Once the boundaries have been detected in each plane, the geometry is formed by interpolating between adjacent cross sections. In a tube-like cavity, similar to Figure 8, where there is only one polygon on each cross section, a simple triangulation is enough to construct the geometry as shown in Figure 8E. However, the vocal tract has a very intricate shape (see Section 4) resulting in cross sections with different number of polygons. Therefore, we also need to address the branching problem (Section 3.3.1), to allow for triangulation (Section 3.3.2).

#### 3.3.1 | Branching problem

The branching problem arises when the topology between two adjacent cross sections changes, which may occur in a variety of ways: a cavity opens, closes, merges, and branches, or a hole appears or disappears. To solve the branching, we basically need to find the correspondence between polygons of neighboring cross sections.<sup>51</sup> To do so, all polygons on a cross section are partitioned according to the topology of the neighboring cross sections. The outcome of the partitioning process is a set of polygons on adjacent cross sections with similar topology and one-by-one correspondence that can be connected without confusion.

Figure 9 illustrates the partitioning process with a few examples. The first row of each example represents the polygons of three consecutive cross sections  $P_{i-1}$ ,  $P_i$ , and  $P_{i+1}$ . Applying the partitioning process on cross section  $i$ ,  $P_i$  is partitioned into three sets of polygons  $C_i^p$ ,  $R_i$ , and  $C_i^n$ , as shown in the second row of each example in Figure 9.  $C_i^p$  and  $C_i^n$  are respectively used to connect the cross section  $i$  to the previous ( $i - 1$ ) and the next ( $i + 1$ ) cross section. Parts of  $P_i$  that are not



**FIGURE 9** Applying partitioning process to a few examples.  $P_i$  polygons on cross section  $i$ ,  $C_i^p$  partitioned set according to the previous cross section,  $R_i$  partitioned residual set, and  $C_i^n$  partitioned set according to the next cross section. The dashed and solid lines in the second row of each example correspond respectively to the  $P_i$  and the partitioning set

shared by  $C_i^p$  and  $C_i^n$  are considered as a residual  $R_i$ , which only relies on the cross section  $i$ . The pseudo code of the partitioning process (see the appendix) consists of two main steps: the PARTSIM and the RESIDUAL. PARTSIM deals with the cross sections in pairs and tries to find regions with similar topology to the previous or the next cross section.  $C_i^p$  and  $C_i^n$  are respectively the outcome of the PARTSIM for polygons  $(P_i, P_{i-1})$  and  $(P_i, P_{i+1})$ . The RESIDUAL routine is used to calculate  $R_i$ .



## PARTSIM

The main idea of the partitioning method is to split a reference polygon  $P_{\text{ref}}$  in order to identify a region  $P_{\text{sim}}$  that has a similar topology to a target polygon  $P_{\text{tar}}$ . To do so, the algorithm starts with the overlapping region ( $P_{\text{sim}} = P_{\text{ref}} \cap P_{\text{tar}}$ ). Then, the distance between vertices of  $P_{\text{sim}}$  and  $P_{\text{tar}}$  and their gradient is calculated. A vertex of  $P_{\text{sim}}$  is replaced by the corresponding vertex of  $P_{\text{tar}}$  if the angle between gradient vectors is less than  $\pi/2$  and the distance between them is less than a threshold. This is to avoid small slivers that may appear when there are small differences between  $P_{\text{ref}}$  and  $P_{\text{tar}}$ . A dissimilar polygon  $P_{\text{dis}}$  is calculated as the difference between  $P_{\text{tar}}$  and  $P_{\text{sim}}$  ( $P_{\text{dis}} = P_{\text{tar}} - P_{\text{sim}}$ ). Parts of  $P_{\text{dis}}$  may join with  $P_{\text{sim}}$  if  $P_{\text{sim}}$  keeps its original topology afterwards.

## RESIDUAL

Once  $C_i^p$  and  $C_i^n$  have been determined by PARTSIM, the residual  $R_i$  is calculated as the symmetric difference between sets  $C_i^p \Delta C_i^n$ . A residual polygon appears when a new cavity opens or an existing cavity closes. They may also appear when the corresponding polygon in the neighboring cross section changes its shape significantly. Some polygons in the residual set may be too small and could be joined with  $C_i^p$  and  $C_i^n$ . The MERGE routine checks the possibility of joining. Merging is only allowed when the topology does not change afterwards. The merged polygon should keep the number of simple polygons and their orientation (clockwise or counterclockwise). It should also keep the bounding box and area ratios of the original and merged polygons below a threshold (20% change is allowed).

The first example (see Figure 9A) illustrates a branching situation. In this example,  $P_{i-1}$  has three polygons. Two of them merge on the next cross section, and  $P_i$  has two polygons. Another merge happens in cross section  $(i + 1)$ , and  $P_{i+1}$  consists of only one polygon. Applying the partitioning process on cross section  $i$  results in three partitioning sets  $C_i^p$ ,  $R_i$ , and  $C_i^n$  as shown in the second row of Figure 9A.  $C_i^p$  and  $P_{i-1}$  are similar in topology while this is not the case for  $C_i^n$  and  $P_{i+1}$ . It is important to note that we should expect similarity in topology between  $C_i^n$  and  $C_{i+1}^p$  and between  $C_i^p$  and  $C_{i-1}^n$ . In this example (Figure 9A), the residual  $R_i$  does not join with  $C_i^p$ , since the topology of  $C_i^p$  would change. After partitioning the cross section  $i$  in Figure 9A, the algorithm repeats the process for the cross section  $i + 1$ , which has been illustrated in Figure 9B. We should expect  $C_i^n$  to be similar to  $C_{i+1}^p$ . This is indeed confirmed in Figure 9B. The third example (see Figure 9C) presents another situation when the shape changes significantly. The algorithm is capable of finding similar regions. The residual  $R_j$  does not merge with  $C_j^p$ , as the bounding box and area would change significantly. The last example (see Figure 9D) depicts a more complex situation where a complex polygon with a hole inside morphs into two simple polygons without any hole. Partitioning sets are identified properly. In this example, merging the residual  $R_k$  with  $C_k^n$  does not change the bounding box at all and may only change the area slightly. However, the orientation of the hole polygon changes. Hence, the algorithm does not consider this merging and keeps the residual.

### 3.3.2 | Triangulation

The partitioning process explained in Section 3.3.1 finds the one-to-one correspondence between polygons of neighboring cross sections.  $C_i^p$  is connected to  $C_{i-1}^n$ ,  $C_i^n$  is connected to  $C_{i+1}^p$ , etc, using a set of triangles. One vertex of each triangle belongs to  $C_{i-1}^n$ , and the other vertex belongs to  $C_i^p$ , and the third vertex belongs to either  $C_{i-1}^n$  or  $C_i^p$ . Figure 8E depicts a simple example where both polygons have approximately the same perimeter and the same number of vertices. In this simple case, the triangles are formed by drawing edges between corresponding vertices. However, in general, the polygons may have different perimeters or different number of vertices. Then a Delaunay triangulation algorithm is used to create the triangles between two consecutive cross sections. Since the residual polygon  $R_i$  has no correspondence, the polygon  $R_i$  is triangulated so that to form a planar surface. The right-most column of Figure 9 shows examples of triangulation.

## 4 | VOCAL TRACT GEOMETRY

The vocal tract is a cavity with a complex shape composed of one main cavity and some other small cavities. The size and shape of these cavities change significantly, and the movement of the structures, in particular the tongue, causes some cavities to appear or disappear. The main cavity includes the larynx tube, laryngopharynx, oropharynx, and oral regions. Other cavities are the piriform fossa, the sublingual, the interdental spaces, and the vallecula. The piriform fossa is formed by the larynx wall, thyroid bone, and pharynx wall. The sublingual cavity is a space between the tongue apex and mouth floor, which is formed when the tongue tip is raised. The interdental cavity is the space between the upper and lower teeth

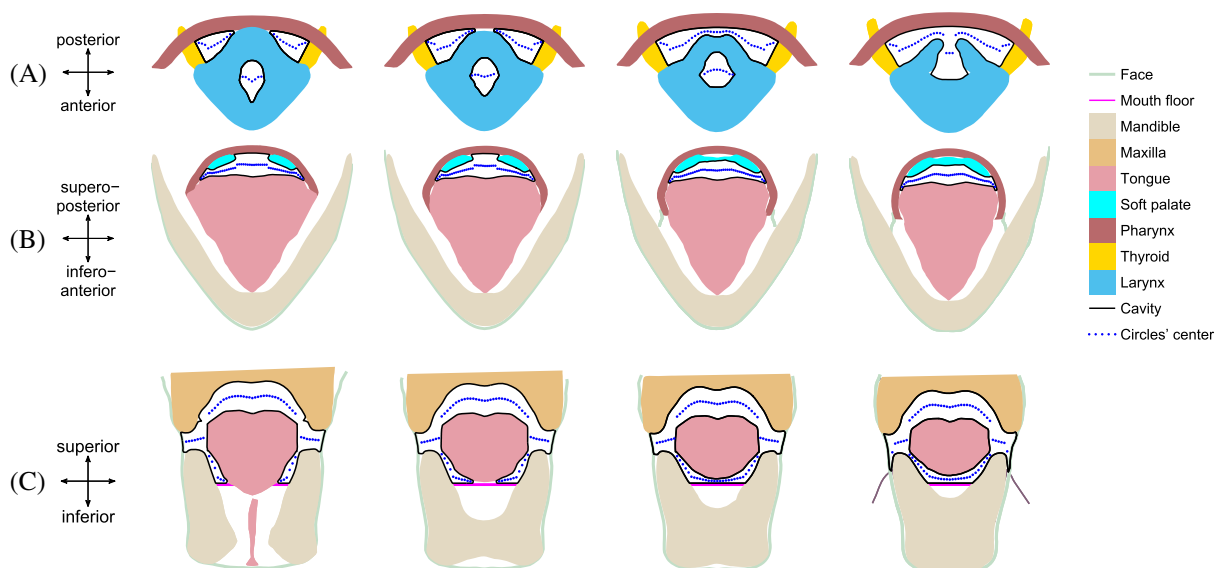
and constitutes a part of the main cavity when the mouth is open. However, raising the tongue detaches this space from the main cavity in some regions. The vallecula, which is another small cavity formed by the tongue root and the epiglottis, is omitted in this work, since the epiglottis is not included in our model. The cavity reconstruction method, explained in Section 3, is applied to reconstruct the vocal tract geometry. The parameter  $K$ , defined in Equation 1, is constant for all of the cross sections. An adequate accuracy is achieved by choosing a value between 1.2 and 1.5.

#### 4.1 | The neutral vocal tract

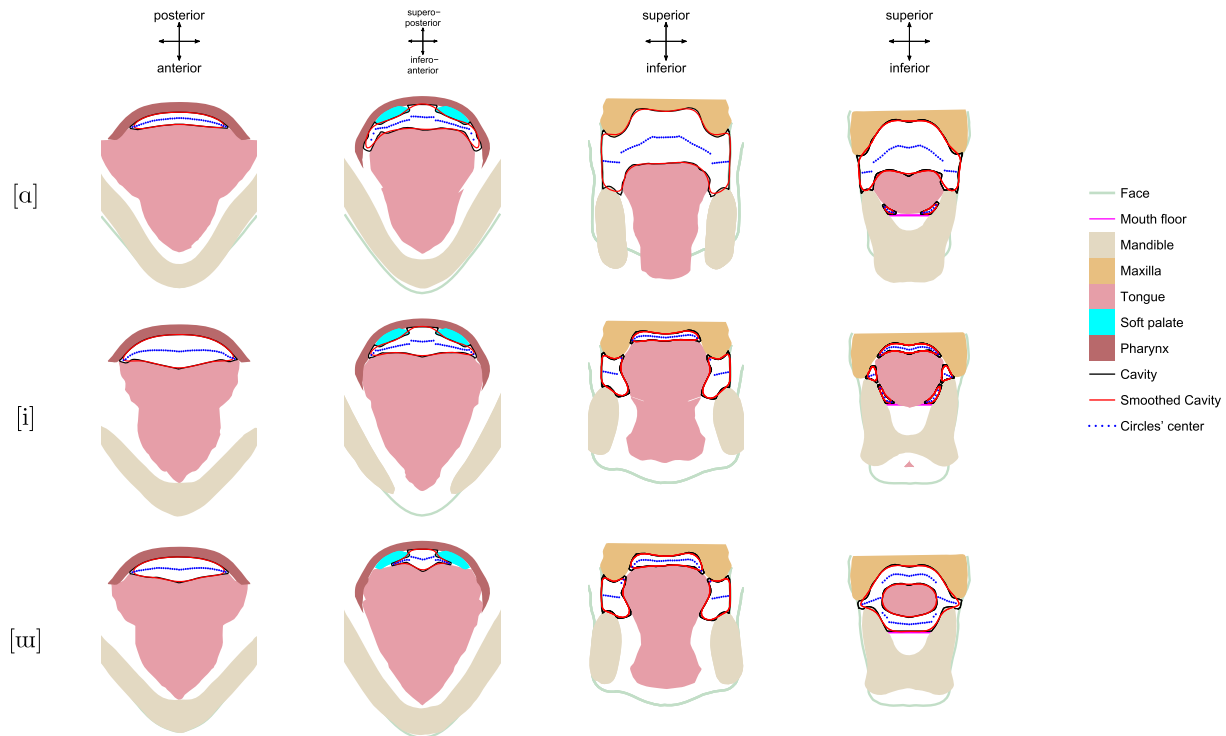
When all muscles are at rest (activations are zero), the biomechanical model is in its neutral position, roughly corresponding to the schwa articulation. To acquire the corresponding vocal tract geometry, the cavity reconstruction method, described in Section 3, is applied. In a similar way as the example in Figure 8, the geometries of the structures are intersected with slicing planes to detect the 2D boundary of the vocal tract. Figure 10 shows a few 2D cross sections and the detected boundaries of the vocal tract. Figure 10A depicts four consecutive cross sections in the laryngopharynx region. As shown in the left-most example of Figure 10A, three polygons corresponding to the left and right piriform fossa and the larynx tube are detected as boundaries of the vocal tract. Other cross sections of Figure 10A show how the left/right piriform fossa merge with the larynx tube. Figure 10B shows cross sections in the velar region with the vocal tract boundaries detected. Figure 10C shows cross sections in the oral cavity, indicating that the interdental spaces and the sublingual cavity have been detected in all cross sections. In the right-most example of Figure 10C, the algorithm considers the lower lip as the boundary of the cavity that results in sharp corners (the area between the lower teeth and face). We address this issue, which may also arise in other cross sections, either in the preprocessing or the postprocessing. In the preprocessing, both upper and lower teeth are identified; a horizontal segment line connects the surface of the teeth to the face to prevent the circles to grow into this small area. If bypassing this step, these sharp corners get smoothed in the postprocessing by applying a Savitzky-Golay filter.<sup>52</sup> The polygons of all cavities are transformed back to 3D, and the shape is then reconstructed. To get a smooth shape, a Laplacian smoother<sup>53</sup> is applied to the 3D shape of the cavity. The reconstructed geometry of the schwa ( $[\text{ə}]$ ) is shown in Figure 12, illustrating that the main cavity, piriform fossa, interdental and sublingual cavities have been reconstructed.

#### 4.2 | Other static geometries

To reconstruct the vocal tract geometry of a given phoneme, the corresponding muscle activations is needed as input to the biomechanical model. However, such activations are not known to a large extent. Electromyography (EMG), as the



**FIGURE 10** Examples of boundary detection in a region A, where the larynx and pharynx tube merge, B, around the soft palate, C, in the oral part with a sublingual cavity



**FIGURE 11** Examples of cross sections and detected boundaries of the vocal tract while producing the three cardinal vowels [ɑ, i, u]

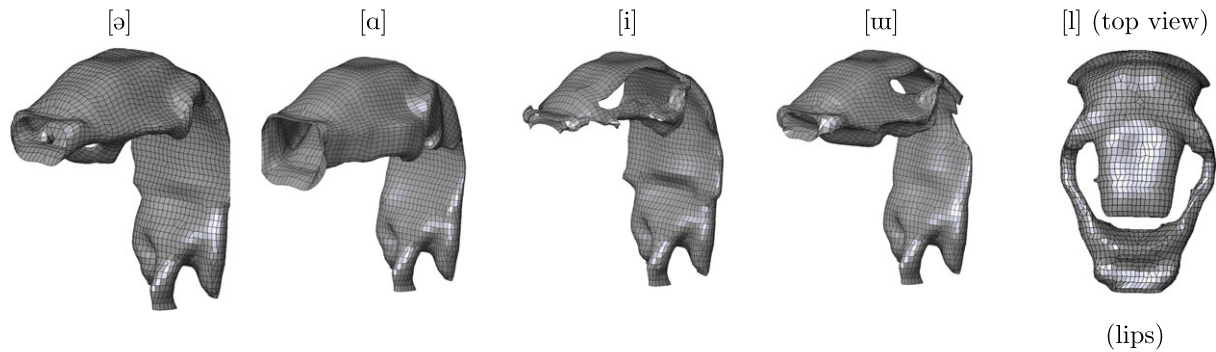
**TABLE 1** Activation of the tongue muscles: zero and hundred represent, respectively, no active and maximum active force generated by the muscle

Phoneme	GGP	GGM	GGA	SG	HG	MH	GH	V	T	IL	SL	JO	JC
ɑ	0	0	25	5	40	0	10	0	0	0	0	10	0
i	100	20	50	20	0	100	50	0	100	0	10	0	2
u	20	0	0	20	7	50	0	0	0	0	0	0	0.5
l	100	0	0	0	0	0	50	0	0	0	30	0	1.5

For abbreviation of the tongue muscles, see Figure 2; JO and JC are two groups of muscles that are responsible for opening and closing the jaw, respectively.

common technique for measuring the muscle activity, is too invasive to be used in speech tasks. For the purpose of this work, we are instead looking for any plausible solution that can generate known articulation to test the cavity reconstruction method. To this end, we use previously published EMG measurements<sup>54</sup> as the initial point for our simulations. The initial activation for a muscle was calculated by normalizing the measurements in Baer et al<sup>54</sup> against the maximum measurement of that muscle across all utterances. In the next step to modify the initial values, we established two criteria, which were used to the largest possible extent: first, to use only the active muscles reported in the measurement and second, to keep the order of activation magnitude the same as in the measurement. For example, if the measurement shows the maximum activity of GGP for vowel [i], then we keep it maximum in our simulations for that vowel. Then the activations were modified until we get a typical tongue shape (see, eg, Stone and Lundberg<sup>55</sup> and Engwall<sup>56</sup>), midsagittal contour of the vocal tract, and the first-two formants<sup>57,58</sup> of the corresponding phonemes. To determine the formant frequencies in this stage, the area function is calculated after reaching the equilibrium (using the method proposed in Dabbaghchian et al<sup>50</sup>), followed by 1D acoustic solver.<sup>59,60</sup> The final activations used in this study are reported in Table 1. Three cardinal vowels were chosen to account for the maximum deformation of the vocal tract. In addition, a lateral approximant was chosen to illustrate how the method deals with complex cavity formed when the oral cavity is blocked in the middle and the air propagates along two lateral channels.

For each phoneme, the corresponding muscle activations (in Table 1) were fed as input to the biomechanical model. After reaching equilibrium, the vocal tract geometry was reconstructed using all geometries of the anatomical structures.



**FIGURE 12** The vocal tract geometries of schwa [ə], three cardinal vowels [ɑ,i,u], and a lateral approximant [l]

Figure 11 illustrates a few examples of cross sections in the pharyngeal and oral cavities for three vowels. In all cases, the vocal tract boundaries are accurately detected. Figure 11 also shows how the tongue shape forms the interdental spaces and the sublingual cavity. The reconstructed geometries of schwa, three vowels, and one lateral approximant are depicted in Figure 12. In all geometries, the piriform fossa and the larynx tube are identical since the larynx structure is static in the biomechanical model. In the vowel [ɑ], the interdental space is part of the main cavity while it gets split in both [i] and [u], because of the raised tongue. This cavity appears as a side branch, which then merges into the main cavity. In vowel [u], the interdental cavity is smaller and merges into the main cavity in a posterior position compared with vowel [i], because of the tongue's posterior placement. This also creates a relatively large sublingual cavity in [u], whereas it is very small in the two other vowels. The lateral approximant [l] is articulated in a different way as the tongue tip touches the palate and blocks the airway in the middle (eg, see, Zhou et al<sup>61</sup> for more details). This branches the airway in the oral cavity into two lateral channels. Figure 12 shows the top view of [l] geometry (this view is chosen to show the lateral channels).

### 4.3 | Dynamic geometries

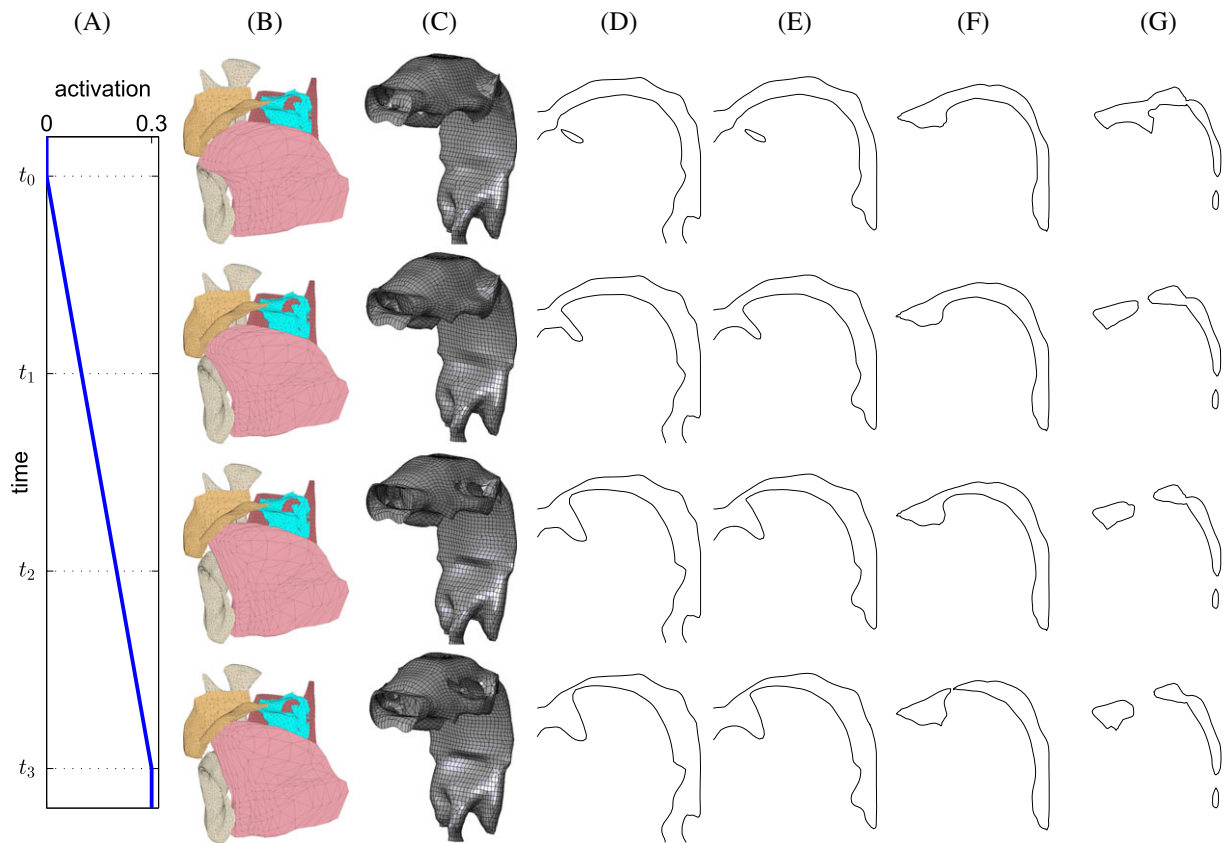
The exact same methodology is used for time-changing vocal tract geometries. As the input to the biomechanical model is a temporal signal of muscles' activation, the geometry of the vocal tract is deforming continuously. Applying the cavity reconstruction at each time step generates a sequence of deforming geometries, as exemplified in Figure 13. In this example, which is chosen to illustrate changes in the sublingual and the interdental cavities, activation of two muscles including GGP and SL is linearly increased from 0 to 0.3. This cause the tongue apex to raise and touch the palate. Since it might be difficult to see differences between 3D geometries, sagittal slices of the geometries are also presented.

## 5 | ACOUSTIC SIMULATION OF VOWELS

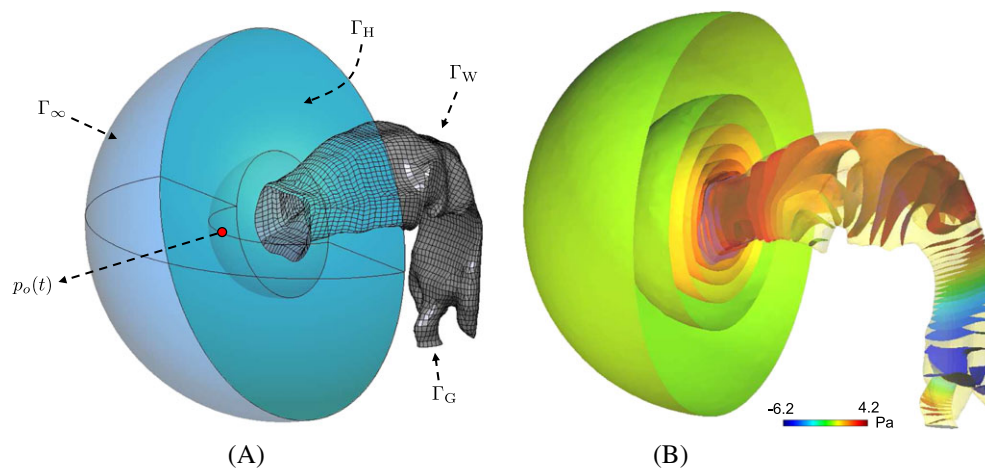
To check the viability of the reconstructed vocal tracts for voice generation, we performed acoustic simulations of the three cardinal vowels. This served as a first proof of concept to validate that the presented approach resulted in vocal tracts suitable for acoustics and flow numerical simulations.

### 5.1 | Method

The wave equation in mixed form for the acoustic pressure and acoustic particle velocity was solved, using the stabilized FEM of Codina<sup>62</sup> in a computational domain consisting of the vowel vocal tract plus a hemisphere of radius 0.08 m. The hemisphere let acoustic waves propagate outwards from the mouth (see Figure 14A). It is to be noted that herein we are only concerned with the generation of static vowel sounds. The production of vowel-vowel utterances would imply setting the mixed wave equation in an arbitrary Lagrangian-Eulerian (ALE) framework, as in Guasch et al.<sup>63</sup> This is however out of the scope of the current work and has been left for future developments linking biomechanical models and acoustic simulations.



**FIGURE 13** An example showing the tongue movement and the corresponding reconstructed geometries of the vocal tract at different time instants: (A) activation of genioglossus posterior (GGP) and superior longitudinal (SL) muscles, (B) movement of the tongue, (C) the corresponding reconstructed vocal tract geometries, (D) midsagittal slice of the vocal tract, (E-G) sagittal slices positioned at 8, 16, and 20 mm from the midsagittal slice



**FIGURE 14** (A) Vocal tract geometry of vowel [a] attached to a hemisphere that allows sound waves radiate from the mouth.  $\Gamma_G$  stands for the glottal cross-sectional area,  $\Gamma_W$  for the vocal tract walls,  $\Gamma_H$  for a circular flat baffle to be interpreted as the head, and  $\Gamma_\infty$  for a nonreflecting boundary. A red dot indicates the point where the acoustic pressure  $p_o(t)$  is tracked. (B) Snapshot of the acoustic pressure distribution of vowel [a] at time instant  $t = 3$  ms, showing 20 isosurfaces equally distributed in the range between [-6.2, 4.2] Pa

The following boundary conditions were considered for the problem in Figure 14A. An input particle velocity  $g(t)$  (to be detailed below) was prescribed at the glottal cross section  $\Gamma_G$ . Wall losses were introduced at the vocal tract walls  $\Gamma_W$  by means of an impedance,  $Z_w$ , which was set to  $Z_w = 83666 \text{ kg/m}^2\text{s}$  according to Švancara and Horáček.<sup>64</sup> The

flat surface of the hemisphere  $\Gamma_H$ , which roughly accounts for the diffraction of the head (see Arnela et al<sup>65</sup> for better approximations), was assumed acoustically rigid, ie,  $Z_H \rightarrow \infty$ . Finally, the well-known Sommerfeld boundary condition was imposed at the outer boundary  $\Gamma_\infty$  to prevent spurious back reflections from it. This boundary condition is only optimal for waves impinging in the normal direction to the surface, so perfectly matched layers (eg, Takemoto et al<sup>4</sup> and Arnela and Guasch<sup>66</sup>) or infinite elements (eg, Švancara and Horáček<sup>64</sup> and Vampola et al<sup>67</sup>) are usually favored instead. However, in our case, we were mainly dealing with spherical waves emanating from the mouth that mostly impact the boundary in the normal direction. Moreover, it was proved in Espinoza et al<sup>68</sup> that the Sommerfeld condition for the wave equation in mixed form also performs well in much more demanding situations, so that condition sufficed in our case. Besides, and in what concerns the physical parameters of the problem, the speed of sound was set to the standard value in speech production of  $c_0 = 350$  m/s, and the air density was taken as  $\rho_0 = 1.1644$  kg/m<sup>3</sup>.

The air volume contained within the vocal tract geometry was meshed with unstructured tetrahedral elements of size  $h \approx 0.0025$  m, whereas  $h$  was designed to smoothly change from  $h \approx 0.0025$  m in the immediate region close to the mouth to  $h \approx 0.005$  m in the curved surface of the hemisphere. This gave about 140.000 elements in total that guaranteed capturing all wavelengths in the frequency range up to 10 kHz. Each vowel vocal tract geometry was characterized by means of its vocal tract transfer function, defined by

$$H(f) = \frac{P_o(f)}{Q(f)}, \quad (2)$$

where  $P_o(f)$  and  $Q(f)$  respectively stand for the Fourier transform of the acoustic pressure  $p_o(t)$  at the exit of the vocal tract and the input volume velocity  $Q(t)$  at the glottal cross section. The latter can be simply related to the acoustic particle velocity  $g(t)$  using the glottal cross-sectional area  $S_g$ ,  $Q(t) = g(t)S_g$ . The Gaussian pulse

$$G_p(t) = e^{-[(t-T)/0.29T]^2} \text{ (m/s)} \quad (3)$$

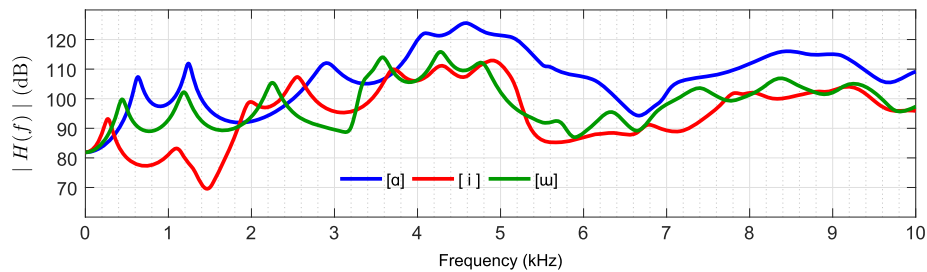
was used for  $g(t)$ , with  $T = 0.646/f_0$  and  $f_0 = 10$  kHz. This pulse was low pass filtered with a cutoff frequency of 10 kHz to avoid the appearance of high frequency numerical errors above the maximum frequency of analysis (10 kHz). A 50-millisecond numerical simulation was then performed using a sampling frequency of  $f_s = 160$  kHz, and the acoustic pressure  $p_o(t)$  was tracked at a node located at 0.03 m in front of the mouth exit (see red dot in Figure 14A).

## 5.2 | Results

The FEM simulations allow one to observe how acoustic waves propagate within the vocal tract and get emitted outwards from the mouth. As an example, in Figure 14B, we show a snapshot of the acoustic pressure distribution for vowel [a] at time instant  $t = 3$  milliseconds. A set of 20 isosurfaces of constant acoustic pressure is represented in the figure to better appreciate the pressure distribution within the vocal tract and outside it. The isosurfaces are equally distributed in the range between  $[-6.2, 4.2]$  Pa, respectively corresponding to the minimum and maximum pressure values at this time instant.

Inside the vocal tract, plane wave propagation mainly occurs in the narrower regions (see, eg, the pharyngeal cavity), while in wider zones, such as the oral cavity, acoustic waves also travel in the transverse direction giving rise to complex 3D pressure distribution patterns, which could not obviously be modeled using 1D approaches. Spherical sound waves radiated from the mouth exit can be also observed in the figure. They reach the outer boundary of the computational domain hemisphere in the normal direction, as mentioned before, which guarantees the correct performance of the Sommerfeld boundary condition.

The computed vocal tract transfer functions  $H(f)$  for each vowel sound are shown in Figure 15. Several resonances can be observed below 4 kHz, which can be attributed to vocal tract eigenmodes (formants in the speech community) associated with plane wave propagation along the vocal tract centerline direction. For all vowels, the location and width of these first formants coincide with those reported in literature (see, eg, Stevens<sup>57</sup> or Arnela et al<sup>69,70</sup> among many others). Beyond this frequency, some spectral dips and further formants also show up, which may be attributed to the presence of side branches like the piriform fossa and also to the generation of higher order modes (see, eg, Takemoto et al<sup>4</sup> and Blandin et al<sup>71</sup>). One can also observe how a formant clustering occurs in the frequency range between 3 and 5 kHz. This is associated



**FIGURE 15** Magnitude of the vocal tract transfer function  $|H(f)|$  for three vowels [a, i, u]

with the singing formant phenomenon that aids projecting the voice when singing and that physically depends on the epilarynx cross-sectional area and length (see, eg, Sundberg<sup>72</sup> and Story<sup>73</sup>). For the vowel [i], a small resonance followed by an antiresonance is produced between 1 and 2 kHz. This is very similar to the observations of Honda et al,<sup>74</sup> where the phenomenon was attributed to the bilateral interdental space. On the other hand, we shall note that none of the vocal tract transfer functions in Figure 15 decay with frequency as observed for the human voice. This is because the definition of  $H(f)$  compensates for this effect. The typical spectral tilt at high frequencies would be recovered once a proper train of glottal pulses was introduced at the glottal cross section.

A thorough analysis of the vocal tract acoustics is out of the scope of this work. However, the above results show that the reconstructed vocal tract from a biomechanical model can, in principle, replicate previous observations of acoustic phenomenon, hence illustrating the large potential of the approach.

## 6 | CONCLUSIONS

The speech production apparatus is composed of several 3D anatomical structures, and a complete modeling of such structures is not possible in lower dimensions than 3D. Figure 2 (showing the fiber direction of the tongue muscles) shows this very clearly. How could the transverse muscle or the partial sagittal spread of the HG be incorporated in a 2D midsagittal model? Also, the airway as the cavity formed by surrounding 3D structures has a complex shape consisting of several cavities. The minor cavities, in addition to the main cavity, may play a role to shape the speech acoustics. In some sounds, these minor cavities even become a major player. For example, the two fricative sounds [s] (eg, “s” in “sip”) and [ʃ] (eg, “sh” in “ship”) are distinguished on the basis of if the sublingual cavity is formed or not.<sup>75,76</sup> In addition, 2D articulatory modeling may not be adequate for some sounds such laterals (eg, [l] in “let”) where the air passes along the lateral sides of the tongue.<sup>61</sup> However, when moving to 3D, the vocal tract geometry may become intricate and/or difficult to determine and dedicated methods are hence required to successfully reconstruct the vocal tract.

In this work, we have proposed a method, namely, cavity reconstruction, that can reconstruct intricate 3D shapes of the airway from biomechanical models. This is achieved by converting the 3D problem into a set of 2D problems, identifying the 2D boundaries, and projecting back into 3D. The idea of the GCs provides an elegant solution to identify the 2D boundaries of the vocal tract despite the existing gaps and overlaps. By addressing the branching problem, the reconstructed geometries include the piriform fossa, interdental space, and sublingual cavity in addition to the main cavity. As a first check of the validity of the reconstructed vocal tracts for acoustics, we have performed FEM simulations for three cardinal vowels. The corresponding vocal tract transfer functions have been computed revealing that the locations and widths of formants are well recovered in all cases.

In this work, we have focused on presenting the cavity construction method and showing its usefulness for a biomechanical model. Note that, since the cavity construction works on geometries of the articulators, it is in fact applicable to all geometrical models, and may even be beneficial for image-based boundary detection, in cases where the resolution or threshold is such that traditional image-based methods lead to null or erroneously large results.

This work may contribute to ambitious computational models that aim to simulate the entire human physiology (eg, see, Hunter et al<sup>77</sup> and Magnenat-Thalmann et al<sup>78</sup>). In particular, it may advance the upper airway modeling for speech production, aspiration, and swallowing applications and may lead to studies that are either difficult to do with human subjects or difficult to capture with imaging technologies such as MRI. As a follow-up study, we will study the acoustic coupling for time-changing vocal tract shapes and investigate the contribution of different cavities in generating acoustic response and their interaction.

## ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 308874. The second and fourth authors would also like to acknowledge the support provided by the Agencia Estatal de Investigación and FEDER, EU, through Project GENIOVOX TEC2016-81107-P.

## ORCID

Saeed Dabbaghchian  <http://orcid.org/0000-0002-8991-1016>

## REFERENCES

- Motoki K. Three-dimensional acoustic field in vocal-tract. *Acoust Sci Technol*. 2002;23(4):207-212. <https://doi.org/10.1250/ast.23.207>
- Zhou X, Espy-Wilson CY, Boyce S, Tiede M, Holland C, Choe A. A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *J Acoust Soc Am*. 2008;123(6):4466-4481. <https://doi.org/10.1121/1.2902168>
- Vampola T, Horáček J, Švec JG. FE modeling of human vocal tract acoustics. Part 1: production of Czech vowels. *Acta Acust United with Acust*. 2008;94(3):433-447. <https://doi.org/10.3813/AAA.918051>
- Takemoto H, Mokhtari P, Kitamura T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *J Acoust Soc Am*. 2010;128(6):3724-3738. <https://doi.org/10.1121/1.3502470>
- Aalto D, Aaltonen O, Happonen RP, et al. Large scale data acquisition of simultaneous MRI and speech. *Appl Acoust*. 2014;83:64-75. <https://doi.org/10.1016/j.apacoust.2014.03.003>
- Speed M, Murphy D, Howard D. Modeling the vocal tract transfer function using a 3D digital waveguide mesh. *IEEE/ACM Trans Audio, Speech Lang Process*. 2014;22(2):453-464. <https://doi.org/10.1109/TASLP.2013.2294579>
- Traser L, Birkholz P, Flügge TV, et al. Relevance of the implementation of teeth in three-dimensional vocal tract models. *J Speech Lang Hear Res*. 2017;60(9):1. [https://doi.org/10.1044/2017\\_JSLHR-S-16-0395](https://doi.org/10.1044/2017_JSLHR-S-16-0395)
- Xu C, Sin S, McDonough JM, et al. Computational fluid dynamics modeling of the upper airway of children with obstructive sleep apnea syndrome in steady flow. *J Biomech*. 2006;39(11):2043-2054. <https://doi.org/10.1016/j.jbiomech.2005.06.021>
- Jeong S-J, Kim W-S, Sung S-J. Numerical investigation on the flow characteristics and aerodynamic force of the upper airway of patient with obstructive sleep apnea using computational fluid dynamics. *Med Eng Phys*. 2007;29(6):637-651. <https://doi.org/10.1016/j.medengphy.2006.08.017>
- Mylavarapu G, Murugappan S, Mihaescu M, Kalra M, Khosla S, Gutmark E. Validation of computational fluid dynamics methodology used for human upper airway flow simulations. *J Biomech*. 2009;42(10):1553-1559. <https://doi.org/10.1016/j.jbiomech.2009.03.035>
- Zhao M, Barber T, Cistulli PA, Sutherland K, Rosengarten G. Simulation of upper airway occlusion without and with mandibular advancement in obstructive sleep apnea using fluid-structure interaction. *J Biomech*. 2013;46(15):2586-2592. <https://doi.org/10.1016/j.jbiomech.2013.08.010>
- Huang C-J, Huang S-C, White SM, Mallya SM, Eldredge JD. Toward numerical simulations of fluid-structure interactions for investigation of obstructive sleep apnea. *Theor Comput Fluid Dyn*. April 2016;30(1-2):87-104. <http://dx.doi.org/10.1007/s00162-015-0372-7>
- Liu H, Moxness MHS, Prot VE, Skallerud BH. Palatal implant surgery effectiveness in treatment of obstructive sleep apnea: a numerical method with 3D patient-specific geometries. *J Biomech*. 2018;66:86-94. <https://doi.org/10.1016/j.jbiomech.2017.11.006>
- Zheng Z, Liu H, Xu Q, et al. Computational fluid dynamics simulation of the upper airway response to large incisor retraction in adult class I bimaxillary protrusion patients. *Sci Rep*. 2017;7:45706. <http://www.nature.com/articles/srep45706>
- Fu M, Barlaz MS, Holtrop JL, et al. High-frame-rate full-vocal-tract 3D dynamic speech imaging. *Magn Reson Med*. 2017;77(4):1619-1629. <https://doi.org/10.1002/mrm.26248>
- Lingala SG, Zhu Y, Kim Y-C, Toutios A, Narayanan S, Nayak KS. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magn Reson Med*. 2017;77(1):112-125. <https://doi.org/10.1002/mrm.26090>
- Zhou X, Woo J, Stone M, Espy-Wilson CY. Three-dimensional vocal tract modeling of fricatives /s/ and /sh/ for post-glossectomy speakers. *Proc Meetings Acoust*. 2013;19:1-4. <https://doi.org/10.1121/1.4799003>
- Takatsu J, Hanai N, Suzuki H, et al. Phonologic and acoustic analysis of speech following glossectomy and the effect of rehabilitation on speech outcomes. *J Oral Maxillofac Surg*. 2017;75(7):1530-1541. <https://doi.org/10.1016/j.joms.2016.12.004>
- Stavness I, Gick B, Derrick D, Fels S. Biomechanical modeling of English /r/ variants. *J Acoust Soc Am*. 2012;131(5):EL355-60. <https://doi.org/10.1121/1.3695407>
- Harandi NM, Woo J, Stone M, Abugharbieh R, Fels S. Variability in muscle activation of simple speech motions: a biomechanical modeling approach. *J Acoust Soc Am*. 2017;141(4):2579-2590. <https://doi.org/10.1121/1.4978420>
- Sonomura M, Mizunuma H, Numamori T, Michiwaki H, Nishinari K. Numerical simulation of the swallowing of liquid bolus. *J Texture Stud*. 2011;42(3):203-211. <https://doi.org/10.1111/j.1745-4603.2011.00287.x>
- Ho AK, Tsou L, Green S, Fels S. A 3D swallowing simulation using smoothed particle hydrodynamics. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2014;2(4):237-244. <https://doi.org/10.1080/21681163.2013.862862>



23. Farazi MR, Martin-Harris B, Harandi NM, Fels S, Abugharbieh R. A 3D dynamic biomechanical swallowing model for training and diagnosis of dysphagia. In: 12th IEEE International Symposium on Biomedical Imaging; 2015; Brooklyn, NY, USA:1385-1388. <https://doi.org/10.1109/ISBI.2015.7164134>
24. Ilegbusi OJ, Kuruppumullage N, Silverman E, Lewis V, Lehman J, Ruddy BH. Mathematical modelling of tongue deformation during swallow in patients with head and neck cancer. *Math Comput Model Dyn Syst.* 2016;22(6):569-583. <https://doi.org/10.1080/13873954.2016.1220015>
25. Kikuchi T, Michiwaki Y, Koshizuka S, Kamiya T, Toyama Y. Numerical simulation of interaction between organs and food bolus during swallowing and aspiration. *Comput Biol Med.* 2017;80:114-123. <https://doi.org/10.1016/j.compbiomed.2016.11.017>
26. Maeda S. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. *Speech Prod. Speech Model.* Netherlands: Springer; 1990:131-149. [https://doi.org/10.1007/978-94-009-2037-8\\_6](https://doi.org/10.1007/978-94-009-2037-8_6)
27. Heinz JM. On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech. Liege; 1965.
28. Perrier P, Boë L-J, Sock R. Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast modeling the transition with two sets of coefficients. *J Speech Lang Hear Res.* 1992;35(1):53-67. <https://doi.org/10.1044/jshr.3501.53>
29. Soquet A, Lecuit V, Metens T, Demolin D. Mid-sagittal cut to area function transformations: direct measurements of mid-sagittal distance and area with MRI. *Speech Commun.* 2002;36(3):169-180. [https://doi.org/10.1016/S0167-6393\(00\)00084-4](https://doi.org/10.1016/S0167-6393(00)00084-4)
30. Payan Y, Perrier P. Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis. *Speech Commun.* 1997;22(2):185-205. [https://doi.org/10.1016/S0167-6393\(97\)00019-8](https://doi.org/10.1016/S0167-6393(97)00019-8)
31. Story BH, Titze IR, Hoffman EA. Vocal tract area functions from magnetic resonance imaging. *J Acoust Soc Am.* 1996;100(1):537-554. <https://doi.org/10.1121/1.415960>
32. Story BH, Titze IR, Hoffman EA. Vocal tract area functions for an adult female speaker based on volumetric imaging. *J Acoust Soc Am.* 1998;104(1):471-487. <https://doi.org/10.1121/1.423298>
33. Kröger BJ, Winkler R, Mooshammer C, Pompino-Marschall B. Estimation of vocal tract area function from magnetic resonance imaging: preliminary results. In: 5th Semin. Bavaria: Speech Prod.; 2000:333-336.
34. Takemoto H, Honda K, Masaki S, Shimada Y, Fujimoto I. Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *J Acoust Soc Am.* 2006;119(2):1037-1049. <https://doi.org/10.1121/1.2151823>
35. Takemoto H. Morphological analyses of the human tongue musculature for three-dimensional modeling. *J Speech Lang Hear Res.* 2001;44(1):95-107. [https://doi.org/10.1044/1092-4388\(2001\)009](https://doi.org/10.1044/1092-4388(2001)009)
36. Buchaillard S, Perrier P, Payan Y. A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning. *J Acoust Soc Am.* 2009;126(4):2033-2051. <https://doi.org/10.1121/1.3204306>
37. Fang Q, Fujita S, Lu X, Dang J. A model-based investigation of activations of the tongue muscles in vowel production. *Acoust Sci Technol.* 2009;30(4):277-287. <https://doi.org/10.1250/ast.30.277>
38. Kajee Y, Pelteret J-PV, Reddy BD. The biomechanics of the human tongue. *Int J Numer Method Biomed Eng.* April 2013;29(4):492-514. <https://doi.org/10.1002/cnm.2531>
39. Wu X, Dang J, Stavness I. Iterative method to estimate muscle activation with a physiological articulatory model. *Acoust Sci Technol.* 2014;35(4):201-212. <https://doi.org/10.1250/ast.35.201>
40. Pelteret JPV, Reddy BD. Development of a computational biomechanical model of the human upper-airway soft-tissues toward simulating obstructive sleep apnea. *Clin Anat.* 2014;27(2):182-200. <https://doi.org/10.1002/ca.22313>
41. Anderson P, Fels S, Harandi NM, et al. Chapter 20—FRANK: a hybrid 3D biomechanical model of the head and neck. In: Payan Y, Ohayon J, eds. *Biomech. Living Organs.* Oxford: Academic Press; 2017:413-447. <https://doi.org/10.1016/B978-0-12-804009-6.00020-1>
42. Stavness I, Sánchez CA, Lloyd J, et al. Unified skinning of rigid and deformable models for anatomical simulations. In: SIGGRAPH Asia 2014 Technical Briefs. Shenzhen: ACM; 2014:9:1-9:4. <https://doi.org/10.1145/2669024.2669031>
43. Dabbaghchian S, Arnela M, Engwall O, Guasch O, Stavness I, Badin P. Using a biomechanical model and articulatory data for the numerical production of vowels. In: Proc. Interspeech. San Francisco, CA, USA; 2016:3569-3573. <https://doi.org/10.21437/Interspeech.2016-1500>
44. Dabbaghchian S, Arnela M, Engwall O, Guasch O. Synthesis of VV utterances from muscle activation to sound with a 3D model. In: Interspeech. Stockholm, Sweden; 2017:3497-3501. <https://doi.org/10.21437/Interspeech.2017-1614>
45. Lloyd JE, Stavness I, Fels S. ArtiSynth: a fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In: Payan Y, ed. *Soft Tissue Biomech. Model. Comput. Assist. Surg.* Berlin: Springer Berlin Heidelberg; 2012:355-394. [https://doi.org/10.1007/8415\\_2012\\_126](https://doi.org/10.1007/8415_2012_126)
46. Harandi NM, Stavness I, Woo J, Stone M, Abugharbieh R, Fels S. Subject-specific biomechanical modelling of the oropharynx: towards speech production. *Comput Methods Biomech Biomed Eng: Imaging Visualization.* 2017;5(6):416-426. <https://doi.org/10.1080/21681163.2015.1033756>
47. Stavness I, Lloyd JE, Payan Y, Fels S. Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics. *Int J Numer Method Biomed Eng.* 2011;27(3):367-390. <https://doi.org/10.1002/cnm.1423>
48. Vatti BR. A generic solution to polygon clipping. *Commun ACM.* 1992;35(7):56-63. <https://doi.org/10.1145/129902.129906>
49. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell.* 1994;16(6):641-647. <https://doi.org/10.1109/34.295913>
50. Dabbaghchian S, Arnela M, Engwall O. Simplification of vocal tract shapes with different levels of detail. In: 18th Int. Congr. Phonetic Sci.; 2015; Glasgow, Scotland, UK:1-5.
51. Bajaj CL, Coyle EJ, Lin K-N. Arbitrary topology shape reconstruction from planar cross sections. *Graph Model Image Process.* 1996;58(6):524-543. <https://doi.org/10.1006/gmpip.1996.0044>

52. Orfanidis SJ. *Introduction to Signal Processing Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.; 1995.
53. Desbrun M, Meyer M, Schröder P., Barr AH. Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proc. 26th annu. conf. comput. graph. interact. tech., Vol. 33. ACM Press/Addison-Wesley Publishing Co.; 1999; New York, NY, USA: 317-324. <https://doi.org/10.1145/311535.311576>
54. Baer T, Alfonso PJ, Honda K. Electromyography of the tongue muscles during vowels in /əpVp/ environment. *Ann Bull RILP*. 1988;22:7-19.
55. Stone M, Lundberg A. Three-dimensional tongue surface shapes of English consonants and vowels. *J Acoust Soc Am*. 1996;99(6): 3728-3737. <https://doi.org/10.1121/1.414969>
56. Engwall O. Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. *Speech Commun*. 2003;41(2-3): 303-329. [https://doi.org/10.1016/S0167-6393\(02\)00132-2](https://doi.org/10.1016/S0167-6393(02)00132-2)
57. Stevens KN. *Acoustic Phonetics*. Cambridge: The MIT Press; 2000.
58. Peterson GE, Barney HL. Control methods used in a study of the vowels. *J Acoust Soc Am*. 1952;24(2):175-184. <https://doi.org/10.1121/1.1906875>
59. Zhou X, Zhang Z, Espy-Wilson C. VTAR: a Matlab-based computer program for vocal tract acoustic modeling. *J Acoust Soc Am*. 2004;115(5):2543-2543. <https://doi.org/10.1121/1.4783643>
60. Espy-Wilson C, Zhou X, Tiede M, Boyce S. New features in VTAR: a Matlab-based computer program for vocal tract acoustic modeling. *J Acoust Soc Am*. 2007;121(5):3136-3136. <https://doi.org/10.1121/1.4782167>
61. Zhou X, Espy-Wilson CY, Tiede M, Boyce S. An MRI-based articulatory and acoustic study of lateral sound in American English. In: IEEE Int. Conf. Acoust. Speech Signal Process. Dallas: IEEE; 2010:4182-4185. <https://doi.org/10.1121/1.4782167>
62. Codina R. Finite element approximation of the hyperbolic wave equation in mixed form. *Comput Methods Appl Mech Eng*. 2008;197(13-16):1305-1322. <https://doi.org/10.1016/j.cma.2007.11.006>
63. Guasch O, Arnela M, Codina R, Espinoza H. A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production. *Acta Acust united with Acust*. 2016;102(1):94-106. <https://doi.org/10.3813/AAA.918927>
64. Švancara P, Horáček J. Numerical modelling of effect of tonsillectomy on production of Czech vowels. *Acta Acust united with Acust*. 2006;92(5):681-688.
65. Arnela M, Guasch O, Alías F. Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations. *J Acoust Soc Am*. 2013;134(4):2946-2954. <https://doi.org/10.1121/1.4818756>
66. Arnela M, Guasch O. Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method. *J Acoust Soc Am*. 2013;133(6):4197-4209. <https://doi.org/10.1121/1.4803889>
67. Vampola T, Laukkanen A-M, Horáček J, Švec JG. Vocal tract changes caused by phonation into a tube: a case study using computer tomography and finite-element modeling. *J Acoust Soc Am*. 2011;129(1):310-315. <https://doi.org/10.1121/1.3506347>
68. Espinoza H, Codina R, Badia S. A Sommerfeld non-reflecting boundary condition for the wave equation in mixed form. *Comput Methods Appl Mech Eng*. 2014;276:122-148. <https://doi.org/10.1016/j.cma.2014.03.015>
69. Arnela M, Blandin R, Dabbaghchian S, et al. Influence of lips on the production of vowels based on finite element simulations and experiments. *J Acoust Soc Am*. 2016;139(5):2852-2859. <https://doi.org/10.1121/1.4950698>
70. Arnela M, Dabbaghchian S, Blandin R, et al. Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds. *J Acoust Soc Am*. 2016;140(3):1707-1718. <https://doi.org/10.1121/1.4962488>
71. Blandin R, Arnela M, Laboissière R, et al. Effects of higher order propagation modes in vocal tract like geometries. *J Acoust Soc Am*. 2015;137(2):832-843. <https://doi.org/10.1121/1.4906166>
72. Sundberg J. Articulatory interpretation of the "singing formant". *J Acoust Soc Am*. 1974;55(4):838-844. <https://doi.org/10.1121/1.1914609>
73. Story BH. Vowel acoustics for speaking and singing. *Acta Acust United with Acust*. 2004;90(4):629-640.
74. Honda K, Takano S, Takemoto H. Effects of side cavities and tongue stabilization: possible extensions of the quantal theory. *J Phon*. 2010;38(1):33-43. <https://doi.org/10.1016/j.wocn.2008.11.002>
75. Perkell JS, Boyce SE, Stevens KN. Articulatory and acoustic correlates of the [s-š] distinction. *J Acoust Soc Am*. 1979;65(S1):S24-S24. <https://doi.org/10.1121/1.2017169>
76. Perkell JS, Matthies ML, Tiede M, et al. The distinctness of speakers' /s-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *J Speech Lang Hear Res*. 2004;47(6):1259-1269. [https://doi.org/10.1044/1092-4388\(2004\)095](https://doi.org/10.1044/1092-4388(2004)095)
77. Hunter P, Coveney PV, de Bono B, et al. A vision and strategy for the virtual physiological human in 2010 and beyond. *Philos Trans R Soc London A Math Phys Eng Sci*; 368(1920):2595-2614. <https://doi.org/10.1098/rsta.2010.0048>
78. Magnenat-Thalmann N, Ratib O, Choi HF, eds. *3D Multiscale Physiological Human*. London: Springer; 2014. <https://doi.org/10.1007/978-1-4471-6275-9>

**How to cite this article:** Dabbaghchian S, Arnela M, Engwall O, Guasch O. Reconstruction of vocal tract geometries from biomechanical simulations. *Int J Numer Meth Biomed Engng*. 2019;35:e3159. <https://doi.org/10.1002/cnm.3159>

## APPENDIX : PSEUDO CODE OF THE PARTITIONING PROCESS

**Algorithm 1** Partitioning Process

---

```

1:  $C_i^p = \text{PARTSIM}(P_i, P_{i-1})$ 
2:  $C_i^n = \text{PARTSIM}(P_i, P_{i+1})$ 
3:  $C_i^p, C_i^n, R_i = \text{RESIDUAL}(C_i^p, C_i^n)$ 

4: function PARTSIM( $P_{\text{ref}}, P_{\text{tar}}$ )
5:    $P_{\text{sim}} = P_{\text{ref}} \cap P_{\text{tar}}$ 
6:   for all  $v_1 \in P_{\text{tar}}$  do
7:     for all  $v_2 \in P_{\text{sim}}$  do
8:        $\text{dist}(v_1, v_2) = |v_1 - v_2|$ 
9:        $\text{grad}(v_1, v_2) = \langle \nabla P_{\text{tar}}|v_1, \nabla P_{\text{sim}}|v_2 \rangle$ 
10:    end for
11:   end for
12:    $\text{dist}(\text{grad} < 0) = \infty$ 
13:   for all  $v_2 \in P_{\text{sim}}$  do
14:      $d_m = \min_{v_1} \text{dist}$ 
15:     if  $d_m < d_t$  then
16:        $v_2 \leftarrow v_1$ 
17:     end if
18:   end for
19:    $P_{\text{dis}} = P_{\text{tar}} - P_{\text{sim}}$ 
20:    $f = \text{MERGE}(P_{\text{sim}}, P_{\text{dis}})$ 
21:   if  $f$  then
22:      $P_{\text{sim}} = P_{\text{sim}} \cup P_{\text{dis}}$ 
23:   end if
24:   return  $P_{\text{sim}}$ 
25: end function

26: function RESIDUAL( $P_1, P_2$ )
27:    $R = P_1 \Delta P_2$ 
28:    $f_1 = \text{MERGE}(P_1, R)$ 
29:    $f_2 = \text{MERGE}(P_2, R)$ 
30:   if  $f_1 \& f_2$  then
31:      $P_1 = P_1 \cup R$ 
32:      $P_2 = P_2 \cup R$ 
33:      $R = \emptyset$ 
34:   end if
35:   return  $P_1, P_2, R$ 
36: end function

37: function MERGE( $P, R$ )
38:    $P_t = P \cup R$ 
39:    $f = P_t$  topology same as  $P$ 

40:   return  $f$ 
41: end function

```

$\triangleright$  find part of  $P_{\text{ref}}$  which is similar to  $P_{\text{tar}}$   
 $\triangleright$  euclidean distance between vertices  
 $\triangleright \nabla P|v$  : gradient of polygon  $P$  at vertex  $v$   
 $\langle \cdot, \cdot \rangle$  : inner product  
 $\triangleright$  do not consider distance if the inner product of the gradient vectors is negative  
 $\triangleright$  search for the closest vertex  
 $\triangleright d_t$  is the distance threshold. It is calculated as the maximum edge length in both  $P_{\text{ref}}$  and  $P_{\text{tar}}$   
 $\triangleright$  replace vertex  
 $\triangleright$  relative complement

$\triangleright$  find residual  
 $\triangleright \Delta$  : symmetric difference

$\triangleright$  check residual polygon merging  
 $\triangleright \cup$  : union  
 $\triangleright$  Two complex polygons have the same topology if they have composed of:

1. the same number simple polygons
2. with the same orientation

and also with:

3. Small difference in their bounding box and area